DOCUMENT RESUME

ED 344 933 TM 018 282

AUTHOR Beller, Michael
TITLE Tree Representation of Similarity and Dominance Relations between Tests and Items.
PUB DATE Apr 92
NOTE 26p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 20-24, 1992).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS College Entrance Examinations; *Decision Making; Difficulty Level; Equations (Mathematics); Foreign Countries; Higher Education; *Hypothesis Testing; Item Response Theory; *Mathematical Models; *Test Items
IDENTIFIERS Additive Trees; Dominance Hierarchies; Hierarchical Decision Models; Israel; Local Independence (Tests); *Preference Trees; *Similarity (Concept)

ABSTRACT

It has previously been shown by M. Beller (1990) that an additive tree (Addtree, a hierarchical tree representation of similarity data developed by S. Sattath and A. Tversky in 1977), may be useful for representing the structure between tests and items through the similarity among them as measured by their intercorrelations. In this study, the preference tree (Pretree, a hierarchical tree representation of preferences among alternatives developed by A. Tversky and S. Sattath in 1979) is proposed as suitable for representing dominance relations among test items, as measured by their relative difficulties. The key assumption underneath the Pretree model tested by the trinary condition will serve as a means for testing hypotheses about the similarity structure among the items. The dominance relation is examined under the scalogram model, the local independence assumption, and item response theory models. Data from the Psychometric Entrance Test administered to applicants to the Hebrew University in Jerusalem (Israel) provide an illustration. Pretree suggests a parsimonious way of representing sets of items with different degrees of dependencies among them. The tree structure will represent the dissimilarity among items. The length of branches will reflect the difficulty level of each item and the ratio of these branches will reflect the degree of dominance between the items. One table, multiple figures, and 19 references are included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
MICHAL BELLER
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED344933

TREE REPRESENTATION OF SIMILARITY AND DOMINANCE RELATIONS BETWEEN TESTS AND ITEMS

Michal Beller

National Institute for Testing and Evaluation

Jerusalem, Israel

A paper presented at the annual meeting of the American Educational Research Association, San-Francisco, April, 1992.

TM018282

BEST COPY AVAILABLE

Research in the area of psychological testing has traditionally made extensive use of representation models (e.g., Cattell, 1971; Davison, 1985; Guilford, 1967; Guttman, 1968; Horn, 1968; Schlesinger and Guttman, 1969; Spearman, 1927; Tryon and Bailey, 1970). Such representations are commonly based on observed correlations among tests. All through history, attempts were made to explain the common variance of test scores. Intercorrelation matrices of mental and ability tests served as input for research, development and review of diverse theories about the structure of human intelligence and the relations among intellectual components. Correlation matrices tend to be very dense and difficult to interpret, and therefore several methods have been suggested to represent them in simpler and more parsimonious ways. These methods have been used to enhance the understanding of the the structure underlying test performance and to corroborate theories on the structure of intelligence. The geometric approach has been used almost exclusively to represent the interrelations among tests; either in the form of the popular factor analysis (e.g., Guilford, 1967; Harman, 1976) or by use of multidimensional scaling techniques (e.g., Davison, 1985; Schlesinger and Guttman, 1969). It should be noted that, despite their differences, factor analytic techniques and MDS models can be classified as geometrical representation models, because they both embed the objects in some coordinate space, so that the distances between the points (defined differently by the different techniques) reflect the similarity between the objects (in this case the correlations among tests, sub-tests or items).

In a previous work (Beller, 1990), it was shown that an additive tree (Addtree, an hierarchical tree representation of similarity data, developed by Sattath and Tversky, 1977) may be useful for representing the structure among tests and items, via the similarity among them (as measured by their intercorrelations). In the analyses of tests and items it was found that Addtree had an advantage over planes (generated by SSA-I, Guttman, 1968), in the majority of the data sets that were analyzed, in terms of goodness-of-fit (both linear and monotone measures). In addition, the grouping of the tests and items into clusters and sub-clusters, as suggested by the tree representation, was straightforward. The interpretation of the data, suggested by the tree, as well as its high fit to the data, led to the conclusion that tests and items belong to the group of stimuli, that are better characterized by a large number of discrete features, rather than a small number of continuous dimensions.

Addtree, developed by Sattath & Tversky (1977), is an hierarchical clustering model, that represents a proximity/distance matrix as an additive or "path length" tree. The distance between the objects is given by the length of the shortest path that joins them along the tree branches. For technical reasons, the tree is displayed in a rectangular shape, and the vertical lines are included only for graphical convenience. To enhance the interpretation of the structure, Addtree selects a root for the tree on the basis of the data. However, the distances in an additive tree do not depend on the choice of the root. A rooted tree representation can be viewed as an hierarchy of clusters, and it can be interpreted in terms of common and distinctive features (along the contrast model, suggested by Tversky, 1977). Under such an

interpretation, the length of each arc of the tree corresponds to the measure of the common features shared by the cluster of all the objects originating from that arc - and by them alone. The terminal arcs of the tree correspond to the measure of the unique features of the respective objects. The horizontal path-length distance between any two objects is the measure of the sum of their distinctive features. Note that features common to all objects are not represented in the tree.

Figure 1 demonstrates an Addtree representation of a set of 42 test-items, derived from the Psychometric Entrance Test (PET). PET is a scholastic aptitude test, used in the admissions procedure of applicants to Israeli universities. This particular set consists of items that are relatively heterogenous in content and of an intermediate level of difficulty (see, Beller, 1990). The Addtree solution suggests a clear hierarchical structure of the items. The fit of the tree to the data (gamma correlations among the items) is high (R^2_l=0.89;R^2_m=0.90).

- Insert Figure 1 about here -

In the present study the preference tree (Pretree, an hierarchical tree representation of preferences among alternatives, developed by Tversky and Sattath, 1979) is suggested as suitable for representing dominance relations among test items, as measured by their relative difficulty. The key assumption underneath the Pretree model, tested by the trinary condition, will serve as means for testing hypotheses about the similarity structure among the items.

The Preference Tree Model (Pretree)

Tversky and Sattath (1979) investigated the representation of choice alternatives as a treelike graph called Pretree. Pretree is a probabalistic choice model that describes choice as a covert hierarchical elimination process. Each alternative is represented as a measurable collection of aspects, and the entire ensemble of aspects is assumed to have a tree structure. At each stage one selects an aspect (i.e., a branch of the tree) with probability that is proportional to its measure and eliminates all the alternatives not belonging to the selected branch. The process continues until a single alternative remains. Each tree branch represents aspects that are common to all alternatives below the level of that branch.

An illustrative example of a Pretree representation (from Tversky and Sattath, 1979) of the choice between food entrees (menu) is presented in Figure 2.

- Insert Figure 2 about here -

Thus, in the above figure, the link labeled λ represents the aspects shared by all Meat entrees but not Fish, θ represents the aspects shared by Steak and Roast Beef but not Lamb or Fish, and γ represents the unique aspects of Lamb.

The similarity between the alternatives is reflected by the general tree structure (i.e. its clusters and sub-clusters). The probability, p(xy), of choosing alternative x (e.g. Steak) over y (e.g. Roast Beef) is equal

to: $\alpha/\beta+\alpha$ which is the ratio between the length of x's terminal-node and the path length between x and y on the Pretree. The degree of preference -

$R(xy) = p(xy)/p(yx) = \alpha/\beta$ - is represented by the ratio of the length of the terminal nodes of x and y respectively.

Tversky and Sattath (1979) claimed that choice behavior appears to be context dependent. That is, the strength of preference of one alternative over another depends on the context of the other available alternatives. Furthermore, choice probability depends not only on the values of the alternatives but also on their similarity or comparability (Tversky, 1972).

It was shown by Tversky and Sattath (1979) that any three elements that form a similarity structure resembling a tree (xy)z -

satisfy the following trinary condition -

If - $p(xy) \geqslant p(yx)$; then - $p(xy)p(yz)p(zx) \geqslant p(xz)p(zy)p(yx)$

Where p(xy) is the probability of choosing x over y.

The trinary condition reflects the similarity hypothesis in that the commonality between alternatives enhances their discriminability. The trinary condition is used to test the key assumption, underlying the Pretree model, that the probability of choosing x over y is more extreme (close to 0 or 1) the larger the similarity among the alternatives is.

The Constant Ratio Model (CRM)

The Constant Ratio model (CRM, Luce 1959) is an alternative choice model, which suggests that when all choice probabilities are nonzero, the ratio $p(x,A)/p(y,A)$ is a constant that depend on x and y but not on the offered set A. Under the binary formulation of CRM -

$$p(xy) = v(x) \;/\; [\; v(x)+v(y) \;] = 1 \;/\; [\; 1 + v(y)/v(x) \;]$$

where v is a nonnegative scale defined on the total set of alternatives.

Tree representation of any set of alternatives (i.e. three) that satisfies the CRM model will be star shaped (as shown in the figure below), since each alternative will be represented entirely by aspects that are unique to it and whose measure is proportional to its scale value under CRM. It has been shown by Luce (1959) that the binary CRM is equivalent to the following product rule -

$$p(xy)p(yz)p(zx) = p(xz)p(zy)p(yx)$$

Testing Structural Hypotheses

Tversky and Sattath (1979) have shown that the trinary condition provides necessary and sufficient condition for the representation of binary choice probabilities as a preference tree. For error free data, therefore, this condition can readily be applied to find a tree structure that is compatible with the data. Alternatively, one can rely on independent judgements (e.g. similarity) for the construction of the tree structure and employ standard iterative maximization methods to estimate its parameters. To evaluate goodness-of-fit, the tree model can be tested

assuming the hypothesized tree structure against the binary version of Luce's (1959) Constant Ratio model (CRM).

At the present study, the preference tree (Pretree) will be applied to represent dominance relations among test-items, as measured by their relative difficulty. The dominance relation of item i over j - D(ij) - is suggested here as an analogue to the probability of choosing alternative x over y - p(xy).

The Dominance Relation Between Test-Items

Let the joint distribution of item i and j be as follows -

i \ j	1	0	
1	a	b	a+b
0	c	d	c+d
	a+c	b+d	N

The dominance of item i over j - D(ij) - is defined as the probability to answer correctly item i, given that only one of the two items (i,j) was answered correctly.

$$D(ij) = p(i=1/i+j=1) = b/b+c$$

respectively -

$$D(ji) = p(j=1/i+j=1) = c/b+c$$

$$0.00 \leqslant D(ij) \leqslant 1.00$$

Where -

D(ij) = 1.00 - i is fully dominant over j

D(ij) = 0.00 - j is fully dominant over i

D(ij) = 0.50 - neither item is dominant over the other (occurs when p(i=1)=p(j=1))

D(ij) is a skew-symetric measure: D(ij) + D(ji) = 1.00

The dominance relation, D(ij), defined above, is a measure of the relative difficulty of item i over j (where P(i=1) and P(j=1) are measures of the absolute difficulty of item i and j respectively).

For a given pair of absolute difficulty levels, there could be various relative difficulties. The more similar the items (as measured by their covariance) the more extreme (close to 0 or 1) the dominance relation among them (holding absolute difficulty level constant).

For example -

i \ j	1	0	
1	60	0	60
0	20	20	40
	80	20	100

D(ij) = 0.00

D(ji) = 1.00

i \ j	1	0	
1	48	12	60
0	32	8	40
	80	20	100

D(ij) = 0.27

D(ji) = 0.73

The Dominance Relation Under Several Psychometric Models

a. The dominance relation under the scalogram model (Guttman's scale) -

Items that constitute a scalogram (Guttman, 1950) has the following joint distribution -

i \ j	1	0	
1	a	b	
0	0	d	

D(ij) = b/b+0 = 1.00

D(ji) = 0/b+0 = 0.00

Thus, under a scalogram, D(ij) equals 1.00, for any pair of items for which p(i=1) > p(j=1).

b. The dominance relation under the local independence assumption -

$$D_\theta(ij) = p_\theta(i=1/i+j=1) = p_\theta(i=1 \cap j=0)/p_\theta(i+j=1) =$$

$$= p_0(i=1 \cap j=0)/p_\theta(i=1 \cap j=0)+p_\theta(i=0 \cap j=1) =$$

$$= p_\theta(i=1)p_\theta(j=0)/[p_\theta(i=1)p_\theta(j=0)+p_\theta(i=0)p_\theta(j=1)] =$$

$$= 1 \;/\; [\; 1 + p_\theta(i=0)p_\theta(j=1)/p_\theta(i=1)p_\theta(j=0) \;] \qquad [1]$$

c. The dominance relation under IRT models (Lord, 1980) -

By substituting p_θ, in equation [1] above, with the respective IRT function, the following results are obtained:

Let - $Ai = ai(\theta-bi)$ and $Aj = aj(\theta-bj)$

c_1. The dominance relation under the three parametric IRT model, can be shown to be -

$$D_\theta(ij) = 1 \;/\; [\; 1 + G \;]$$

$$\text{where } G = \frac{[(1-ci)\exp(-Ai)][(1+cj)\exp(-Aj)]}{[(1+ci)\exp(-Ai)][(1-cj)\exp(-Aj)]}$$

c_2. The dominance relation, under the two parametric IRT model, will be -

$$D_\theta(ij) = 1 / [\ 1 + \exp(Aj-Ai)\ ]$$

c_3. The dominance relation, under the one parametric IRT model, will be -

$$D_\theta(ij) = 1 / [\ 1 + \exp(bi-bj)\ ] = D(ij)$$

It can be shown that D(ij), under the one parametric IRT model, is equivalent to p(ij) under the Constant Ratio Model (see also Andrich, 1978 and Brogden, 1977):

Let exp-bj = v(j) and exp-bi = v(i)

$$D(ij) = 1 / [\ 1 + \exp(bi-bj)\ ] = 1 / [\ 1 + \text{exp-bj/exp-bi}\ ] =$$
$$= 1 / [\ 1 + v(j)/v(i)\ ] = p(ij) \quad \text{under CRM}$$

In general, under the above three IRT models, D(ij) is determined by parameters describing each item independently of the others (e.g. difficulty level, discriminability). D(ij) under the one parametric model, as well as under any model for which ai=aj=a, is special in that it is independent of θ. In other words, the dominance of item i over j, under the one parametric IRT model, is constant within each level of ability, as well as across all levels of ability. This feature enables using aggregate data for examining predictions based on the one

parameteric IRT model. For the other two IRT models, on the other hand, Θ should be held constant, and predicitions should be examined within each level of Θ.

Testing Psychometric Structural Hypotheses

The psychometric version of the similarity assumption, that underlies the Pretree model, can be formulated as follows: for any two items, with known levels of difficulty, the dominance relation will be more extreme (close to 0 or 1) the more related these items are.

The trinary condition may be utilized to test hypothesized similarity structures among items. For any three items i,j,k - for which D(ij) > D(ji) and the assumed similarity structure is (ij)k, the following inequality (trinary condition) should hold -

$$D(ij)D(jk)D(ki) > D(ik)D(kj)D(ji)$$

It was shown above that D(ij), under the one parametric IRT model, equals p(ij) under the CRM model. Therefore, for items satisfying the one parametric IRT model, the following equality should hold -

$$D(ij)D(jk)D(ki) = D(ik)D(kj)D(ji)$$

The tree structure of such unidimensional items would look like a fan (star shape).

The trinary condition will be suggested here as a way to test structural hypotheses about the relations among test items. For any given structural hypothesis it will be possible to check whether the trinary condition is satisfied in the direction predicted by the Pretree model. In other words, if the trinary condition is satisfied for a given data set, the similarity structure, against which it was tested, is corroborated.

For example -

Let i and j be 2 items from a Math test, and k an item from a Vocabulary test. A logical hypothesis as to their similarity structure would be of the form (ij)k -

Let the joint distributions among these items be as follows -

i \ j	1	0	
1	55	25	80
0	5	15	20
	60	40	

i \ k	1	0	
1	60	20	80
0	10	10	20
	70	30	

j \ k	1	0	
1	45	15	60
0	25	15	40
	70	30	

If the above structural hypothesis holds, then -

D(ij)D(jk)D(ki) should be larger than D(ik)D(kj)D(ji)

25/30 x 15/40 x 10/30 ? 20/30 x 25/40 x 5/30

3,750/36,000 > 2,500/36,000

The above result corroborates the hypothesized similarity structure, and therefore rejects an alternative unidimensional hypothesis that would have resulted in a degenerate tree like this -

Testing the trinary condition within homogeneous sets of items constitutes, in practice, a test of the unidimensionality assumption underlying this item pool. For the unidimensionality assumption to hold for a certain item pool - the trinary condition should not be satisfied more frequently than chance level.

An Empirical Demonstration

To demonstrate how the trinary condition can be used to examine the structure among items, several sets of items were analyzed. The additive tree (Addtree, Sattath & Tversky, 1977) served as the similarity structure against which the trinary condition was checked.

Method

The data is derived from the Psychometric Entrance Test (PET), administered to applicants to the Hebrew University on 1980. PET is a scholastic aptitude test, comprised of 230 multiple choice items, measuring various cognitive domains.

Fourteen sets of items were selected from the total item pool: 8 sets being homogeneous in content and item-type (i.e., General-Knowledge,

Logic, Math, Numerical Series, Figures, Vocabulary, Analogies and English) and 6 sets were heterogeneous, including items from the various sub-tests. The dominance relation, among each pair of items, and the trinary condition, among the triples, were computed separately for each of the 14 sets of items. The similarity structures, against which the trinary condition was tested, were extracted from Addtree representations of the items within each set (based on the gamma correlations among the items included in each set).

The trinary condition was examined for every triple (i,j,k) with similarity structure (ij)k and for which D(ij) > D(ji).

$$H_o: p\{ D(ij)D(jk)D(ki) > D(ik)D(kj)D(ji) \} = 0.50$$

$$H_1: p\{ D(ij)D(jk)D(ki) > D(ik)D(kj)D(ji) \} > 0.50$$

Let - $d(ijk) = X(ijk) - Y(ijk)$

where - $X(ijk) = D(ij)D(jk)D(ki)$ and $Y(ijk) = D(ik)D(kj)D(ji)$

then -

$$H_o: p\{ \bar{d}(ijk) > 0.00 \} = 0.50$$

$$H_1: p\{ \bar{d}(ijk) > 0.00 \} > 0.50$$

In order to reject H_o in favor of H_1, it is necessary to show that the proportion of cases ($\overline{II}$), for which the trinary condition is satisfied, in the predicted direction, is significantly greater than 0.50. The results are presented in Table 1.

- Insert Table 1 about here -

As an example, let us focus on the third set of items in Table 1. This group consists of 42 items with moderate difficulty level, for which the Addtree representation was previously shown (Fig. 1). The number of triples for which the trinary condition was examined is - N=9,751. The condition was satisfied, in the predicted direction, in $\overline{TT}$=86% of these cases.

No simple statistical test is readily available to test whether the observed proportion $\overline{TT}$ is significantly greater than 0.5, because the various triples are not independent. The dependencies among the triples affect the variance of the statistic $\overline{TT}$. This may be corrected by redefining the degrees of freedom in the X^2 test for differences among proportions. Instead of using variance estimate based on N triples, a smaller sample size, n, was used. This later sample is based on the number of triples for which any pair of items appears only once (independent triples). In the above mentioned example, the degrees of freedom were therefore reduced from N=9,571 to n=357 and X^2=180.56 ($p<.05$). In general it can be seen that in all 14 sets of items $\overline{TT}$ is significantly greater than 0.50 and H_o is rejected on the basis of the corrected X^2 test.

To obtain some indication about the effect size (W), the ratio of the average $\bar{d}(ijk)$ to its Std ($W=\bar{d}/Std$) was computed. Note that under H_o the average W should equal 0. The effect sizes are listed in Table 1 and they tend to be larger for the heterogenous sets of items, although they

are not small even for the homogeneous groups of items.

As was mentioned before, examination of the trinary inequality provides an estimate-free comparison of a degenerate tree structure, that resembles a fan, to the psychometric Pretree model. The results show that in all data sets, the one parametric IRT model is violated in a direction implied by the similarity hypothesis assumed by the tree structure (in this case Addtree). Similar results were derived for the two and three parametric IRT models (holding Θ constant), but the results will not be presented here due to lack of space. In particular, in the case of homogeneous sets of items (e.g. Analogies, Vocabulary, Logic, Numerical Series) H_o is equivalent to the structure derived from the unidimensionality assumption under the IRT models, thus rejecting H_o, in these cases, can be perceived as rejection of the unidimensionality assumption under these models.

Estimating the Length of the Pretree Branches -

Once the trinary condition is satisfied it is possible to estimate the length of the branches in the tree structure against which it was tested. Chandler's (1969) STEPIT program can be employed to search for maximum likelihood estimates of the parameters of the tree, using the observed dominance relations as raw data. The same procedure can be applied under the assumption of a degenerate tree structure, that corresponds to the one parametric IRT model. These two different structural trees can then be compared as to their goodness-of-fit to the data (dominance relations among the items) using X^2.

As an example, the length of the tree-branches were estimated for the Analogies item-pool, both for the case where the structure was derived from the additive tree analysis (i.e. Pretree) and for the degenerate tree structure, which corresponds to the one parametric IRT model. The fit of each tree to the data was computed using X^2:

Pretree	One Parametric IRT model
$X^2 = 4.49$	$X^2 = 26.904$
Df = 86	Df = 91

The difference between the two tree structures, in terms of goodness-of-fit to the data, was found to be significant:

$X^2 = 26.904 - 4.49 = 22.414$ Df = 91-86 = 5 $p < 0.05$

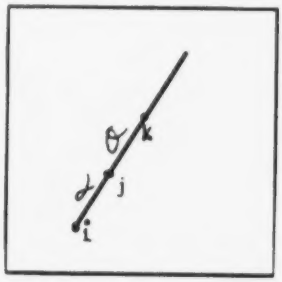
The tree structure implied by the one parametric IRT model was rejected in favor of the more complex Pretree, thus rejecting the unidimensionality assumption underneath this data.

Conclusions

It is concluded that any theory, which attempts to describe the structure of the relations among test items, should take into account parameters relating to the joint distribution of all pairs of items, and not just the parameters of single items (such as their difficulty level, differentiation, etc.).

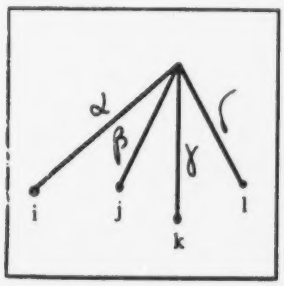
Pretree suggests a parsimonious way of representing sets of items with different degrees of dependencies among them:

a. completely related items - as in Guttman's scale



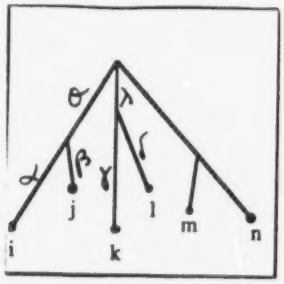
D(ij) =(α)/(α+0) = 1.00 and D(ik) =(α+θ)/(α+θ+0) = 1.00

b. Local independence (as in IRT models) -



D(ij) = $\alpha / \alpha+\beta$ and D(ik) = $\alpha / \alpha+\gamma$

c. Partially related items -



$D(ij) = (\alpha)/(\alpha+\beta)$ and $D(ik) = (\alpha+\theta)/(\alpha+\theta+\lambda+\gamma)$

Under such representation the tree structure will reflect the dissimilarity among the items. The length of the branches will reflect the difficulty level of each item, while the ratio between these branches will reflect the degree of dominance between the items.

REFERENCES

Andrich, D. (1978). Relationship between the Thurstone and Rasch approaches to item scaling. Applied Psychological Measurement, 1978, 2, 449-460.

Beller, M. (1990). Tree versus geometric representation of tests and items. Applied Psychological Measurement, 14(1), 13-28.

Brogden, H. (1977). The Rasch model, the law of comparative judgement and additive conjoint measurement. Psychometrika, 42, 631-634.

Cattell, R. B. (1971). Abilities: Their Structure, Growth and Action. Boston: Houghton-Mifflin.

Chandler, J. P. (1969). Stepit - Finds local minima of a smooth function of several parameters. Behavioral Sciences, 14, 81-82.

Davison, M.L. (1985). Multidimensional scaling versus components analysis of test intercorrelations. Psychological Bulletin, 97, 94-105.

Guilford, J. P. (1967). The Nature of Human Intelligence. New-York: McGraw-Hill.

Guttman, L. (1950). The principal component of scale analysis. In S.S. Stouffer et al. (Eds.) Measurement and Prediction. Princeton, N.J.: Princeton University Press.

Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. Psychometrika, 33, 469-506.

Harman, H. H. (1976). Modern Factor Analysis. Chicago: University of Chicago Press.

Horn, J. L. (1968). Organization of abilities and the development of intelligence. Psychological Review, 75, 242-259.

Lord, F. M. (1980). Application of Item Response Theory to Practical Testing Problems. Hillsdale, N.J.: Erlbaum.

Luce, R. D. (1959). Individual Choice Behavior: A Theoretical Analysis. New-York: Wiley.

Sattath, S. & Tversky, A. (1977). Additive similarity trees. Psychometrika, 42, 319-345.

Schlesinger, I. M., & Guttman, L. (1969). Smallest space analysis of intelligence and achievement tests. Psychological Bulletin, 71, 95-100.

Tryon, T. C., & Bailey, D. E. (1970). Cluster Analysis. New-York: McGraw-Hill.

Tversky, A. (1972). Elimination by aspects: A theory of choice. Psychological Review, 79, 281-299.

Tversky, A. (1977). Features of similarity. Psychological Review, 84, 327-352.

Tversky A. & Sattath, A. (1979). Preference trees. Psychological Review, 86, 542-573.

Table 1

Examination of the trinary condition (TC) within 14 sets of items

Data set	no. items	N	$\overline{\Pi}$	n	x^2	$\bar{d}$	Std	W
Heterogeneous item-sets								
1. random sample	62	36,092	.82	779	31.4*	.007	.009	.78
2. easy items#	42	9,323	.80	321	115.1*	.005	.007	.71
3. intermediate dif#	42	9,751	.86	357	180.6*	.007	.008	.88
4. difficult items#	42	9,310	.83	310	135.0*	.010	.013	.77
5.15 items x GK,V,E	45	13,719	.74	346	82.6*	.006	.009	.67
6 15 items x L,M,N,F	60	20,750	.77	421	125.0*	.006	.009	.67
Homogeneous item-sets								
7. General-Knowl.(GK)	45	11,055	.70	331	53.0*	.004	.007	.57
8. Logic (L)	30	3,272	.67	86	9.9*	.003	.008	.38
9. Math (M)	25	1,823	.81	97	37.3*	.008	.009	.89
10.Numer. Series (N)	20	888	.65	63	5.7*	.004	.010	.40
11.Figures (F)	25	2,042	.84	128	59.2*	.011	.014	.78
12.Vocabulary (V)	35	5,089	.62	203	11.7*	.003	.009	.33
13.Analogies	15	343	.70	33	5.3*	.005	.013	.38
14.English (E)	35	6,252	.61	253	12.3*	.004	.013	.31

* $p<0.05$

6 items from each of the following sub-tests: GK, L, M, N, F, V, E

N - No. of triples for which TC was examined
$\overline{\Pi}$ - proportion of triples for which TC was satisfied in the hypothesized direction
n_2 - no. of triples for which each pair of items appears only once
x^2 - significance test for the difference between $\overline{\Pi}$ and 0.5
$\bar{d}$ - average d(ijk), where d(ijk) = X(ijk) - Y(ijk)
Std - Std of d(ijk)
W - effect size ($\bar{d}$/Std)

Figure 1

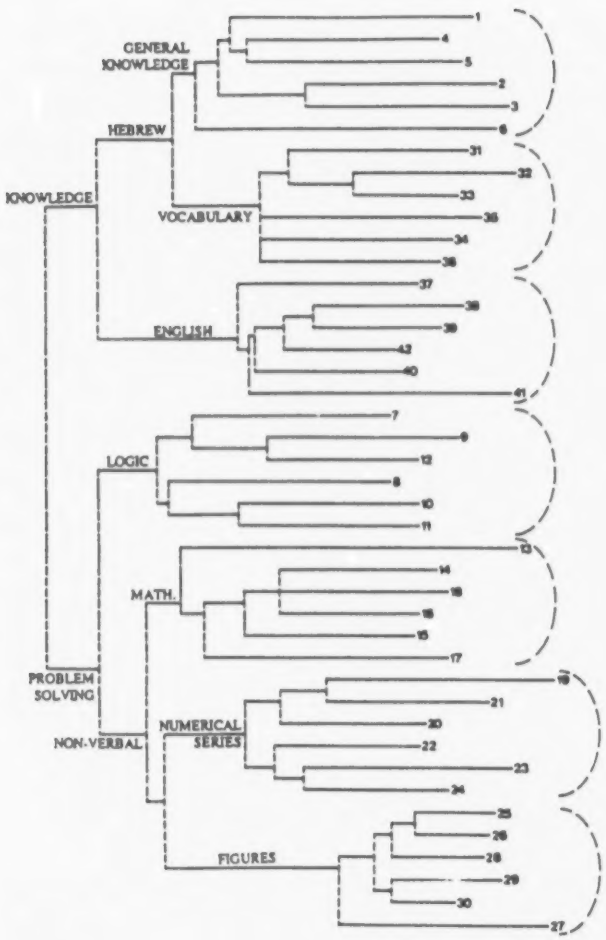


Fig. 2